



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



RDA: punto de vista de usuarios de la comunidad de ciencias de la Tierra

Valencia,
20/09/2018

Pierre-Antoine Bretonnière, Francisco Doblas-Reyes

¿Quiénes somos?

- Departamento de ciencias de la Tierra del BSC (~80 personas)
- Gestionando un catálogo de:
 - ~2PB de datos heterogéneos (csv, txt, grib1/2, netcdf3/4, ...)
 - proveniente de fuentes externas o producidos por nuestros propios modelos
 - de distintas comunidades (meteorología, clima, calidad del aire,...)
 - creciendo a un ritmo de ~75TB/semana
- Con la necesidad de estandarizar los datos para compartirlos con socios, y usar herramientas comunes a la comunidad

RDA en la comunidad de ciencias de la Tierra

- Creación en Marzo del 2016 de un “RDA Interest Group on Weather, climate and air quality”
- 2 workshops organizados por RDA/BSC sobre provenance en ciencias de la tierra
- Muchas temáticas son comunes con otros RDA IG/WG:
 - grandes volúmenes/movimientos de datos
 - gestión de datos muy heterogéneos
 - necesidad de repositorios comunes

RDA en la comunidad de ciencias de la Tierra

- pero hay particularidades:
 - muy avanzado con respecto a estándares de formatos de datos -> la gente puede ver a RDA como demasiado genérico
 - existen infraestructuras internacionales distribuidos (ESGF) pero cuyo acceso puede resultar complicado
 - mucho trabajo se ha hecho con la estandarización de los formatos (grib1/2/3, netcdf3/4), los metadatos y la gobernanza (CF, WMO) pero dejando un poco atrás el “provenance”

¿Cuál es la situación actual de nuestras instituciones estatales ?

- Necesidad de infraestructuras data-centric además de centros de supercomputación (i.e. ICTS)
- La comunidad está dispersa y fragmentada: muchas instituciones generan observaciones y simulaciones (AEMET, IEO, Comunidades Autónomas, ayuntamientos, empresas..) que no están disponibles públicamente, o mantenidas o que no usan estándares

¿Cuál es la situación actual de nuestras instituciones estatales ?

- Se necesita un punto de entrada para acceder a los datos medioambientales españoles que pueda servir:
 - de “broker” a todo lo que esté disponible
 - para guardar y curar datasets almacenados en sistemas pequeños, poco documentados o con pocos recursos (por ejemplo los datos generados en los proyectos financiados por los programas del ministerio de Ciencias).



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



Gracias



pierre-antoine.bretonniere@bsc.es